ELSEVIER
Discrete Applied Mathematics 75 (1997) 231–253
DISCRETE APPLIED MATHEMATICS

On the subword equivalence problem for morphic words

Isabelle Fagnot *
LITP-IBP, Université Paris 6, 2, Place Jussieu, 75252 Paris Cedex 05, France

Received 10 January 1995; revised 17 June 1996

Abstract

Two infinite words x and y are said to be *subword equivalent* if they have the same set of finite subwords (factors). The *subword equivalence problem* is the question whether two infinite words are subword equivalent. We show that, under mild hypotheses, the decidability of the subword equivalence problem implies the decidability of the ω-sequence equivalence problem, a problem which has been shown to be decidable by Čulik and Harju for morphic words (i.e. words generated by iterating a morphism). Yet, we do use the decidability of the ω-sequence equivalence problem to prove our result.

We prove that the subword equivalence problem is decidable for two morphic words, provided the morphisms are primitive and have bounded delays. We also prove that the subword equivalence problem is decidable for any pair of morphic words in the case of a binary alphabet.

Our results hold in fact for a stronger version, namely for the subword inclusion problem.

1. Introduction

The problem we consider here is the subword equivalence problem, that is to say: given two infinite words, is it decidable whether their finite factors are the same?

This problem is of interest in several contexts. First, it is well-known that two infinite words generate the same discrete dynamical system if and only if they have the same set of subwords (see e.g. [10]).

Next, it is easy to show that, under mild hypotheses, the decidability of the subword equivalence problem implies the decidability of the equality. This latter problem remained open for a long time in the case of D0L-systems and has been solved by Čulik and Harju [4].

In this paper, we consider the problem for *morphic* words, that is words obtained by iterating a morphism and we solve it in particular cases.

* Email: Fagnot@litp.ibp.fr.

PII S0166-218X(95)00091-1

We show that the subword equivalence problem is decidable for two morphic words generated by primitive morphisms with bounded delay (Theorem 19 (bis)). As a matter of fact, the proof is by reducing the problem to the ω-sequence equivalence problem and to apply the theorem of Čulik and Harju. It appears that the decidability holds even for the subword inclusion problem.

More generally, we show that the inclusion $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$ is decidable for two morphic words x and y generated by morphisms with bounded delay, if the morphism which generates x is primitive, if the one which generates y is everywhere growing and if x is not ultimately periodic (Theorem 19).

In the case of a binary alphabet, both conditions on the morphisms can be overcome (Theorem 31) by using methods which are standard in the theory of D0L-systems (see e.g. [5]). Consequently, the subword equivalence problem and the subword inclusion problem are decidable for two binary morphic words.

The paper is organized as follows. In Section 2, we provide the main notations and definitions and we recall a few previous results. In Section 3, we prove a 'tiling' lemma and two results upon morphisms with bounded delay. Section 4 is devoted to the demonstration of theorem which deals with the general case. Finally, in Section 5, we consider the binary case.

2. Definitions and notations

Let A be a finite alphabet. The set of finite words is denoted by A^* and the set of infinite words by A^ω. Let f be an endomorphism on A^*. If there exists a letter $a \in A$ such that $f(a) \in aA^*$ and $\lim_{n\to\infty} |f^n(a)| = \infty$ then we can define the infinite word $\mathsf{x} = f^\omega(a)$ as the unique infinite word which has $f^n(a)$ as its prefix for any integer n. Then x is said to be *generated* by iterating f and f is said to be a *generator* of x. We say that an infinite word is *morphic* when it is generated by iterating a morphism (see [1]).

A morphism f on A^* is said to be *nonerasing* if

$$\forall a \in A, \quad |f(a)| \geqslant 1.$$

In the following, we shall consider only nonerasing morphisms.

Given two words u and v, we write $u \leqslant v$ to denote that u is a prefix of v. Given $r \in \mathbb{N}$, we denote $\mathrm{Pref}_r(u)$ the prefix of u of length r if $|u| \geqslant r$; otherwise u. Likewise, we denote $\mathrm{Suff}_r(u)$ the suffix of u of length r, if $|u| \geqslant r$; otherwise u. Given two words u and v on A, we say that u and v are *comparable*, denoted by $u \lessgtr v$, if $u \leqslant v$ or $v \leqslant u$.

Let $\mathsf{x} = a_0 a_1 \cdots a_n \cdots$ be an infinite word over A. For integers $i \leqslant j$, we define $\mathsf{x}[i,j) = a_i a_{i+1} \cdots a_{j-1}$ and $\mathsf{x}[i,j] = a_i a_{i+1} \cdots a_j$. The set of all the finite factors (subwords) of x is denoted by $\mathrm{Fact}(\mathsf{x})$.

$$\mathrm{Fact}(\mathsf{x}) = \{\mathsf{x}[i,j) \mid 0 \leqslant i \leqslant j\}.$$

Lastly, we denote Alph(x) the set of all the letters which have an occurrence in x. We denote σ the *shift operator* on $A^* \cup A^\omega$, defined as follows: if u is a word over A, then $\sigma(u)$ is the unique word such that $u = \mathrm{Pref}_1(u)\sigma(u)$.

An endomorphism f on A^* has *bounded delay* $p \geqslant 1$ *from left to right* if

$$\forall a_1,\dots,a_p,b_1,\dots,b_p \in A, \quad f(a_1\cdots a_p) \leqslant f(b_1\cdots b_p) \Rightarrow a_1 = b_1.$$

It is clear that this condition is equivalent to

$$\forall a_1,\dots,a_p,b_1,\dots,b_p \in A, \quad f(a_1\cdots a_p) \not\approx f(b_1\cdots b_p) \Rightarrow a_1 = b_1.$$

The endomorphism f is said to be *prefix* if it has bounded delay 1, i.e. if

$$\forall a,b \in A, \quad (f(a) \leqslant f(b)) \Rightarrow (a = b).$$

The endomorphism f is said to be *primitive* if there exists an integer n such that

$$\forall a,b \in A, \quad a \in \mathrm{Fact}(f^n(b)).$$

The endomorphism f is said to be *everywhere growing* if

$$\forall a \in A, \quad |f(a)| \geqslant 2.$$

An infinite word x is said to be *recurrent* if any factor of this word has an infinite number of occurrences in x. An infinite word x has *bounded gaps* if for any factor u of x, there exists an integer d such that, for any integer i, the word u has an occurrence in x$[i,i+d)$. It is obvious that, in particular, a word which has bounded gaps is recurrent too. Moreover, a morphic word generated by a primitive morphism has bounded gaps.

An infinite word is said to be *strongly repetitive* if there exists a non-null word u such that, for any integer n, the word u^n is a factor of x (cf. [6]), in this case u is said to be a *near-period* of x.

The following property is classical.

Property 1. *An infinite word* x *is ultimately periodic if and only if there exists an integer n such that there is at most n factors of* x *of length n.*

We also recall this elementary property of which we give a proof in view of reader's convenience.

Property 2. *A non-periodic infinite word with bounded gaps is not strongly repetitive. In particular, a non-periodic morphic word generated by a primitive morphism is not strongly repetitive.*

Proof. We have only to prove the first claim. Let x be a word with bounded gaps which is also strongly repetitive. We are going to show that it is periodic by use of Property 1. Let u be a near-period of x and n be its length. Let v be a factor of length n of x. Since x has bounded gaps, there exists an integer d such that for any integer m, we have $v \in \mathrm{Fact}(\mathsf{x}[m, m+d))$. Furthermore, there exists an integer k such that $|u^k| = kn \geqslant d$ and there exists an integer m_0 such that $\mathsf{x}[m_0, m_0 + kn) = u^k$. Now, v is a factor of $\mathsf{x}[m_0, m_0 + kn) = u^k$, so v is conjugate of u, i.e. there exists an integer ℓ such that $v = \mathrm{Suff}_\ell(u) \cdot \mathrm{Pref}_{n-\ell}(u)$. There is at most n distinct conjugates of u. Then, the word x has at most n factors of length n, so it is ultimately periodic (cf. Property 1) and thereby periodic because it has bounded gaps. □

Now let us consider the following.

Problem 3. *Given two infinite words* x *and* y, *the* subword equivalence problem *is to decide whether they have the same sets of finite factors, i.e., whether* $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{y})$. *The* subword inclusion problem *is to decide if we have* $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$.

Example 4. If the word x is recurrent, then for any suffix y of x, the equality $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{y})$ holds.

Example 5. Let A be the alphabet $\{a, b, c\}$. Let f and g be the endomorphisms on A^* defined by

$$f : \begin{array}{l} a \mapsto aaba \\ b \mapsto ac \\ c \mapsto abc \end{array} \quad \text{and} \quad g : \begin{array}{l} a \mapsto abaa \\ b \mapsto ca \\ c \mapsto bca. \end{array}$$

Let $\mathsf{x} = f^\omega(a)$ and $\mathsf{y} = g^\omega(a)$. Then, on one hand, $g(\mathsf{x}) = a^{-1}\mathsf{x}$ and $f(\mathsf{y}) = a\mathsf{y}$. On the other hand, every factor of x has an infinite number of occurrences in x. Therefore, we can deduce that x and y have the same factors. (A proof of a similar case is given in Lemma 29.)

Example 6. Let f be a primitive morphism. If there exist two distinct letters a and b such that $f(a) \in aA^+$ and $f(b) \in bA^+$, then the words $\mathsf{x} = f^\omega(a)$ and $\mathsf{y} = f^\omega(b)$ have the same factors.

For example, let A be the binary alphabet $\{0, 1\}$. Let f be the endomorphism on A^* defined by

$$f : \begin{array}{l} 0 \mapsto 01 \\ 1 \mapsto 10. \end{array}$$

This morphism generates two infinite words (called Thue–Morse words)

$$\mathsf{t} = f^\omega(0) = 0110100110010110\cdots$$

and

$$\mathsf{t}' = f^\omega(1) = 1001011001101001\cdots.$$

None of the words $\boldsymbol{t}$ and $\boldsymbol{t}'$ is suffix of the other; they even have no common prefix. However, it can easily be seen that they are subword equivalent.

For morphic words, the subword equivalence problem is a generalization of the equality problem.

Given an infinite word x and a letter S, let $\tau(S, \mathsf{x})$ be the infinite word obtained by replacing the first letter of x by S. With these notations, we have the following easy proposition.

Proposition 7. *Let $\mathscr{L}$ be a family of infinite words that are obtained by a given process. Let us suppose that the following three conditions are satisfied*

1. *If $\mathsf{x} \in \mathscr{L}$ and $S \notin \mathrm{alph}(\mathsf{x})$, then $\tau(S, \mathsf{x}) \in \mathscr{L}$;*
2. *The mapping $(S, \mathsf{x}) \mapsto \tau(S, \mathsf{x})$ is effective.*
3. *The mapping $x \mapsto \mathrm{Pref}_1(\mathsf{x})$ is effective.*

If the subword equivalence problem is decidable for $\mathscr{L}$ then the equality problem is decidable for $\mathscr{L}$.

Remark 8. The process may be, for example, a Turing machine, a tag system (cf. [2]) or as in this paper a morphism.

The second condition in the proposition means that from the process defining x, we can effectively deduce the process defining $\mathsf{y} = \tau(S, \mathsf{x})$.

Proof. Let x and y be two infinite words over an alphabet A and S be a letter which does not belong to A. We notice that

$$(\tau(S, \mathsf{x}) = \tau(S, \mathsf{y})) \Leftrightarrow (\mathrm{Fact}(\tau(S, \mathsf{x})) = \mathrm{Fact}(\tau(S, \mathsf{y}))).$$

Indeed, owing to the condition imposed upon S, the set of all prefixes of the word $\tau(S, \mathsf{x})$ is $\mathrm{Fact}(\tau(S, \mathsf{x})) \cap SA^*$. Since, moreover, two infinite words are equal if and only if they have the same prefixes, we obtain the above-mentioned equivalence. Now, it is not difficult to see that

$$\begin{aligned}(\mathsf{x} = \mathsf{y}) &\Leftrightarrow (\mathrm{Pref}_1(\mathsf{x}) = \mathrm{Pref}_1(\mathsf{y}) \text{ and } \tau(S, \mathsf{x}) = \tau(S, \mathsf{y}))\\ &\Leftrightarrow (\mathrm{Pref}_1(\mathsf{x}) = \mathrm{Pref}_1(\mathsf{y}) \text{ and } \mathrm{Fact}(\tau(S, \mathsf{x})) = \mathrm{Fact}(\tau(S, \mathsf{y}))). \qquad \square\end{aligned}$$

Example 9. The family of morphic words satisfies the above conditions. Indeed, let x be a morphic word generated by an endomorphism f on A^*. Let a be a letter such that $\mathsf{x} = f^{\omega}(a)$ and u be a word such that $f(a) = au$. Then,

$$\mathsf{x} = f^{\omega}(a) = a \cdot u \cdot f(u) \cdot f^2(u) \cdots.$$

Let S be a letter which does not belong to A. We define the endomorphism φ on $A \cup \{S\}$ by

$$\varphi(S) = Su, \qquad \varphi(x) = f(x) \quad (x \in A).$$

Then

$$\begin{aligned}\varphi^{\omega}(S) &= S \cdot u \cdot f(u) \cdot f^2(u) \cdots \\ &= \tau(S, \mathsf{x}). \qquad \square\end{aligned}$$

Now let us recall two decidability results which we shall need later. The first one, due to Čulik and Harju, is about the equality of words.

Theorem 10 (Čulik and Harju [4]). *Let* x *and* y *be two morphic words over* A^*. *Then the equality* $\mathsf{x} = \mathsf{y}$ *is decidable.*

The second one, proved independently by, on the one hand, Harju and Linna and, on the other hand, Pansiot, is about the periodicity.

Theorem 11 (Harju and Linna [7], and Pansiot [9]). *Let* z *be a morphic word. We can decide whether* z *is ultimately periodic, and if so, we can effectively compute a preperiod and a period of* z, *i.e. words* u *and* v *such that* $\mathsf{z} = uv^{\omega}$.

3. Preliminary results

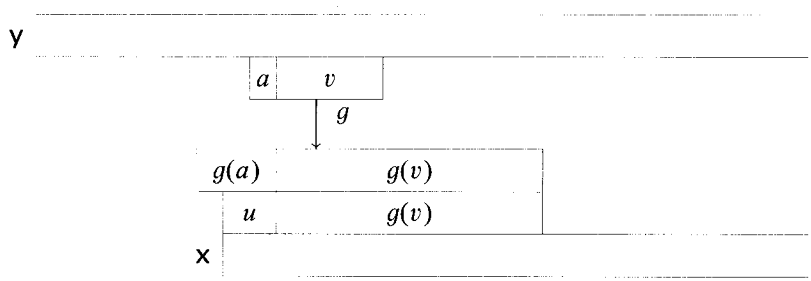
In the following, we shall need a sort of 'tiling' lemma.

Proposition 12. *Let* A *and* B *be two alphabets and* $\mathsf{x} \in B^{\omega}$, $\mathsf{y} \in A^{\omega}$. *Let* g *be a nonerasing morphism* $A^* \to B^*$. *If* $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(g(\mathsf{y}))$, *then there exists* $\mathsf{z} \in A^{\omega}$ *such that* x *is a suffix of* $g(\mathsf{z})$ *and* $\mathrm{Fact}(\mathsf{z}) \subseteq \mathrm{Fact}(\mathsf{y})$.

We also have $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(g(\mathsf{z})) \subseteq \mathrm{Fact}(g(\mathsf{y}))$.

Proof. For any letter a of A and for any word u of B^* such that u is a suffix of $g(a)$, we define the set $\mathscr{B}_{u,a}$ made up of all the words v such that av is a factor of y and such that $ug(v)$ is a prefix of x:

$$\mathscr{B}_{u,a} = \{v \in A^* \mid av \in \mathrm{Fact}(\mathsf{y}) \text{ and } ug(v) \leqslant \mathsf{x}\},$$



The $\mathscr{B}_{u,a}$ are prefix-closed sets, i.e. $\forall v \in \mathscr{B}_{u,a}$, $\forall w \leqslant v$, $w \in \mathscr{B}_{u,a}$.

The union of the $\mathscr{B}_{u,a}$ is infinite. Indeed, since $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(g(\mathsf{y}))$, for any sufficiently long prefix w of x (of length $\geqslant 2\max_{a\in A} g(a)$), there exist two letters a, b in A, a word v in A^* and two integers r and s such that $w = \mathrm{Suff}_r(g(a)) \cdot g(v) \cdot \mathrm{Pref}_s(g(b))$ and $avb \in \mathrm{Fact}(\mathsf{y})$. Then $v \in \mathscr{B}_{\mathrm{Suff}_r(g(a)),a}$. It is not difficult to see that there exists an infinite number of such v. Therefore, $\cup\mathscr{B}_{u,a}$ is infinite.

Moreover, there is a finite number of $\mathscr{B}_{u,a}$. Therefore there exist u_0 and a_0 such that $\mathscr{B}_{u_0,a_0}$ is infinite. Let r be an integer such that $u_0 = \sigma^r(g(a_0))$. Since $\mathscr{B}_{u_0,a_0}$ is prefix-closed, there exists an infinite word $\mathsf{t} \in A^\omega$ such that for any prefix v of t, $v \in \mathscr{B}_{u_0,a_0}$. Then $u_0 g(v) \leqslant \mathsf{x}$; hence, $u_0 g(\mathsf{t}) \leqslant \mathsf{x}$, which is equivalent to $u_0 g(\mathsf{t}) = \mathsf{x}$. Now, let $\mathsf{z} = a\mathsf{t}$. Then on the one hand, $\sigma^r(g(\mathsf{z})) = \mathsf{x}$, so x is a suffix of $g(\mathsf{z})$, and on the other, for any prefix u of z, the word u is a factor of y. Thus $\mathrm{Fact}(\mathsf{z}) \subseteq \mathrm{Fact}(\mathsf{y})$. □

Moreover, we need also two technical results about morphisms with bounded delay.

Proposition 13. *Let f be an everywhere growing endomorphism. If f has bounded delay p then for any integer n, the morphism f^n has bounded delay $2p-1$.*

If $p = 1$, i.e. if f is prefix, we do not need to suppose that f is everywhere growing. Before giving the proof, we need the following lemma.

Lemma 14. *Let $f_1 : B^* \to C^*$ and $f_2 : A^* \to B^*$ be two morphisms. If f_1 has bounded delay $\pi_1 = 2\rho + 1$, $\rho \in \mathbb{N}$ and if f_2 is everywhere growing and has bounded delay π_2, then the morphism $f_1 \circ f_2$ has bounded delay $\rho + \pi_2$.*

In the general case, i.e. if the morphism f_2 is not everywhere growing, we know that $f_1 \circ f_2$ has bounded delay $\pi_1 + \pi_2 - 1$. We draw inspiration from the proof of this result given in [11, Lemma 4.8, p. 67].

Proof. Let $a_1, \ldots, a_{\rho+\pi_2}, b_1, \ldots, b_{\rho+\pi_2} \in A^*$. Let us assume that

$$f_1 \circ f_2(a_1 \cdots a_{\rho+\pi_2}) \Leftrightarrow f_1 \circ f_2(b_1 \cdots b_{\rho+\pi_2}). \tag{1}$$

We want to show that in this case $a_1 = b_1$. To this end, let $u = f_2(a_1 \cdots a_{\pi_2})$, $v = f_2(a_{\pi_2+1} \cdots a_{\rho+\pi_2})$, $u' = f_2(b_1 \cdots b_{\pi_2})$ and $v' = f_2(b_{\pi_2+1} \cdots b_{\rho+\pi_2})$. Then, Eq. (1) can be written

$$f_1(uv) \Leftrightarrow f_1(u'v').$$

Since f_2 is everywhere growing, $|v| \geqslant 2\rho = \pi_1 - 1$ and $|v'| \geqslant 2\rho = \pi_1 - 1$. As a result, the fact that f_1 has bounded delay π_1 implies that $u \Leftrightarrow u'$. That is to say

$$f_2(a_1 \cdots a_{\pi_2}) \Leftrightarrow f_2(b_1 \cdots b_{\pi_2}),$$

and since f_2 has bounded delay π_2, we have $a_1 = b_1$. □

Proof of Proposition 13. We show the claim by induction on n. When $n = 1$, it is obvious by noticing that $p \leqslant 2p - 1$. Next, if we assume that f^n has bounded delay

$2p-1$, we have only to apply the previous lemma with $f_1 = f^n$, $f_2 = f$, $\rho = p-1$ and $\pi_2 = p$ to deduce that f^{n+1} has bounded delay $p+(p-1) = 2p-1$. $\square$

We need to construct the following morphisms.

Definition 15. For all integers m, we define the alphabet

$$B_m = \left\{ \begin{bmatrix} a_1 a_2 \cdots a_m \\ b_1 b_2 \cdots b_m \end{bmatrix} \in A^m \times A^m \mid \forall i, 1 \leqslant i \leqslant m, a_i, b_i \in A, a_1 \neq b_1 \right\}.$$

Given two words $u, v \in A^*$ and an integer m, we define the word $u/_m v$ as follows: if there exist words $w, s, t \in A^*$ and letters $a_1, \ldots, a_m, b_1, \ldots, b_m, a_1 \neq b_1$ such that

$$u = wa_1 \cdots a_m s, \qquad v = wb_1 \cdots b_m t,$$

then

$$u/_m v = \begin{bmatrix} a_1 \cdots a_m \\ b_1 \cdots b_m \end{bmatrix} s;$$

otherwise $u/_m v = 1$.

Let f be an everywhere growing endomorphism over A^* which has bounded delay p. Let $q = 4p-3$. We define the endomorphism $\varphi = T_q(f)$ over $(A \cup B_q)^*$ by

$$\varphi(a) = f(a) \quad (a \in A),$$
$$\varphi \begin{bmatrix} x \\ y \end{bmatrix} = f(x)/_q f(y) \quad \left(\begin{bmatrix} x \\ y \end{bmatrix} \in B_q \right).$$

The endomorphism φ has the following interesting property.

Proposition 16. *For any integer n and any letter $\begin{bmatrix} x \\ y \end{bmatrix} \in B_q$, we have*

$$\varphi^n \begin{bmatrix} x \\ y \end{bmatrix} = f^n(x)/_q f^n(y).$$

i.e.

$$T_q(f^n) = (T_q(f))^n.$$

Remark 17. Again, if f is prefix, we do not need to assume that f is everywhere growing. This remark will be useful when we shall study the case of the binary alphabet.

Proof. To begin with, let us notice that for any integer $n \neq 0$, the fact that f^n is everywhere growing and has bounded delay $2p-1$ makes sure that for any letter $\begin{bmatrix} a_1 a_2 \cdots a_q \\ b_1 b_2 \cdots b_q \end{bmatrix} \in B_q$,

$$f^n(a_1 a_2 \cdots a_q)/_q f^n(b_1 b_2 \cdots b_q) \neq 1.$$

Indeed, f^n has bounded delay $2p-1$, so there exist words u, v, v' such that

$$f^n(a_1a_2\cdots a_{2p-1}) = uv, \qquad f^n(b_1b_2\cdots b_{2p-1}) = uv',$$

and $\mathrm{Pref}_1(v) \neq \mathrm{Pref}_1(v')$.

Moreover, let $w = f^n(a_{2p}\cdots a_q)$, and $w' = f^n(b_{2p}\cdots b_q)$. Since f is everywhere growing, we have $|w| \geqslant 2(2p-2) = 4p-4$ and $|w'| \geqslant 4p-4$. So,

$$f^n(a_1a_2\cdots a_q) = uvw, \qquad f^n(b_1b_2\cdots b_q) = uv'w',$$

with $|vw| \geqslant 4p-3 = q$, $|v'w'| \geqslant q$ and $\mathrm{Pref}_1(vw) \neq \mathrm{Pref}_1(v'w')$, which implies that $f^n(a_1a_2\cdots a_q)/_q f^n(b_1b_2\cdots b_q) \neq 1$.

We prove the stated property by induction on n: For $n=0$, it is obvious. For $n=1$, it is also true by definition of φ. Now, let us assume that the property is true at step n. Let $\left[\begin{smallmatrix}a\\b\end{smallmatrix}\right] \in B_q$. Since f^n has bounded delay $2p-1$, there exist $p,s,t \in A^*$, $\left[\begin{smallmatrix}c\\d\end{smallmatrix}\right] \in B_q$ such that

$$f^n(a) = pcs, \qquad f^n(b) = pdt \tag{2}$$

(cf. the above-mentioned remark). By induction

$$\varphi^n \begin{bmatrix}a\\b\end{bmatrix} = \begin{bmatrix}c\\d\end{bmatrix} s. \tag{3}$$

Moreover, there exist $p', s', t' \in A^*$, $\left[\begin{smallmatrix}c'\\d'\end{smallmatrix}\right] \in B_q$ such that

$$f(c) = p'c's', \qquad f(d) = p'd't', \tag{4}$$

and, so by construction of φ

$$\varphi \begin{bmatrix}c\\d\end{bmatrix} = \begin{bmatrix}c'\\d'\end{bmatrix} s'. \tag{5}$$

But

$$\left.\begin{aligned} f^{n+1}(a) &= f(p)p'c's'f(s)\\ f^{n+1}(b) &= f(p)p'd't'f(t)\end{aligned}\right\} \quad \text{by (2) and (4)}$$

and

$$\varphi^{n+1}\begin{bmatrix}a\\b\end{bmatrix} = \begin{bmatrix}c'\\d'\end{bmatrix} s'f(s) \quad \text{by (3) and (5).}$$

So, the property is also true at step $n+1$. $\square$

This result can be illustrated by Fig. 1.

Example 18. For the morphism of bounded delay 2 (here $q=5$)

$$f: \begin{aligned} a &\mapsto ab\\ b &\mapsto aba,\end{aligned}$$

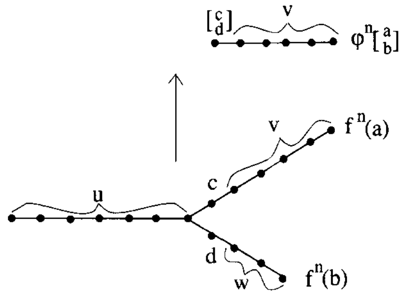


Fig. 1. here $p = 1$ i. e. f is prefix, $q = 1$

we have

$$f(aabba) = aba\,.\,babaa\,.\,baab,$$
$$f(bbaba) = aba\,.\,abaab\,.\,abaab.$$

So,

$$\varphi\begin{bmatrix} aabba \\ bbaba \end{bmatrix} = \begin{bmatrix} babaa \\ abaab \end{bmatrix} baab.$$

Moreover,

$$f^2(aabba) = ababaababa\,.\,ababa\,.\,abababaababababa,$$
$$f^2(bbaba) = ababaababa\,.\,baaba\,.\,babaababaabababa,$$

and, indeed we obtain

$$\begin{aligned}\varphi^2\begin{bmatrix} aabba \\ bbaba \end{bmatrix} &= \varphi\left(\begin{bmatrix} babaa \\ abaab \end{bmatrix} baab\right)\\ &= \begin{bmatrix} ababa \\ baaba \end{bmatrix} abababaabababa\\ &= f^2(aabba)/_q f^2(bbaba).\end{aligned}$$

4. Main result

The main result is the following theorem (Theorem 19) of which we also give a few variants.

Theorem 19. *Let* x *and* y *be two morphic words generated by morphisms with bounded delay such that the morphism which generates* x *is primitive and the one which generates* y *is everywhere growing. If* x *is not ultimately periodic, then the inclusion* $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$ *is decidable.*

Remark 20. Presently, we are not able to prove that the inclusion $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$ is decidable if the word x is ultimately periodic. However, it seems reasonable to think it is the case.

Theorem 19 (bis). *Let x and y be two morphic words generated by primitive morphisms with bounded delay. Then the inclusion $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$ is decidable.*

Theorem 19 (ter). *Let x and y be two morphic words generated by morphisms with bounded delay such that the morphism which generates x is primitive and the one which generates y is everywhere growing. Then the equality $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{y})$ is decidable.*

We shall first show how these variants follow from Theorem 19. Next, we shall prove Theorem 19.

Proof of Theorem 19 (bis). We use the following algorithm of decidability.

Algorithm 1. At first, we test whether x is ultimately periodic, which is possible in view of Theorem 11

(i) If x is ultimately periodic, we test whether y is ultimately periodic;
  (a) if y is not ultimately periodic, then we cannot have $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$. Indeed, the morphism which generates y is primitive and so y cannot be strongly repetitive (see Property 2);
  (b) otherwise, if y is ultimately periodic: we can compute the periods and the preperiods of x and y by means of Theorem 11. Then we can easily decide whether $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$;

(ii) if neither x nor y is ultimately periodic, we apply Theorem 19 in order to decide whether $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$. □

Proof of Theorem 19 (ter). Let f be a primitive morphism with bounded delay which generates x and g be an everywhere growing morphism with bounded delay which generates y. To begin with, let us notice that

(i) If $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{y})$, then the fact that f is primitive and g is everywhere growing implies that g is primitive.

(ii) If $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{y})$ and if x is ultimately periodic, then y too is ultimately periodic.

Therefore, the following algorithm is used.

Algorithm 2. At first, we test whether x is ultimately periodic, which is possible in view of Theorem 11.

(i) If x is ultimately periodic, we test whether y is ultimately periodic;
  (a) if y is not ultimately periodic, then we cannot have $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{y})$ (cf. Remark ii).

(b) otherwise, if y is ultimately periodic, we can compute the periods and the preperiods of x and y by means of Theorem 11. Then we can easily decide whether $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{y})$;

(ii) if neither x nor y is ultimately periodic, we can decide whether $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$ (Theorem 19). Futhermore, by Remark i, g is actually primitive and since f is primitive a power of it is everywhere growing, we can suppose that f is every growing. So we can also decide whether $\mathrm{Fact}(\mathsf{y}) \subseteq \mathrm{Fact}(\mathsf{x})$ by Theorem 19.

In fact, since the set $\mathscr{C}$ of Theorem 19 is symmetric, we have to apply the algorithm one time only. The primitiveness of f and g makes sure that in case of positive answer $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{y})$.

Now, let us give the proof of the main theorem.

Proof of Theorem 19. We are going to show that we can construct a finite set $\mathscr{C}$ of couples of morphic words such that we have the inclusion $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$ if and only if there exists a couple $(\mathsf{x}', \mathsf{y}')$ in $\mathscr{C}$ such that $\mathsf{x}' = \mathsf{y}'$. In order to decide the inclusion $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$, we shall only have to test the equality $\mathsf{x}' = \mathsf{y}'$ for all couples $(\mathsf{x}', \mathsf{y}')$ of $\mathscr{C}$, by means of Theorem 10.

Let f be a primitive morphism with bounded delay which generates x and g be an everywhere growing morphism with bounded delay which generates y. Since f is primitive, there exists an integer n such that f^n is everywhere growing. Consequently, in the following, we shall assume that f is everywhere growing. Let p be such that f and g have bounded delay p. Let us denote $q = 4p - 3$, $\varphi = T_q(f)$ and $\psi = T_q(g)$.

To begin with, let us construct the set $\mathscr{C}$: Whenever there exist letters $\left[\begin{smallmatrix}\mu\\ \nu\end{smallmatrix}\right] \in B_q$, $\alpha \in A$, words $s, s' \in A^*$ and an integer p such that the following three conditions are satisfied:

$$\mu\alpha \in \mathrm{Fact}(\mathsf{x}),$$

$$\varphi^p\left(\begin{bmatrix}\mu\\ \nu\end{bmatrix}\alpha\right) = \begin{bmatrix}\mu\\ \nu\end{bmatrix}\alpha s,$$

$$\psi^p\left(\begin{bmatrix}\mu\\ \nu\end{bmatrix}\alpha\right) = \begin{bmatrix}\mu\\ \nu\end{bmatrix}\alpha s',$$

p is minimal for these two latter properties.

we define the endomorphisms φ' and ψ' over $(A \cup \{S\})^*$, where S is a new letter, by

$$\varphi' : \begin{array}{l} S \mapsto Ss \\ x \mapsto \varphi^p(x) = f^p(x) \quad (x \in A), \end{array}$$

and

$$\psi' : \begin{array}{l} S \mapsto Ss' \\ x \mapsto \psi^p(x) = g^p(x) \quad (x \in A). \end{array}$$

The set $\mathscr{C}$ will be the set of all couples $(\varphi'^{\omega}(S), \psi'^{\omega}(S))$. It is not difficult to see that this set is finite and constructible.

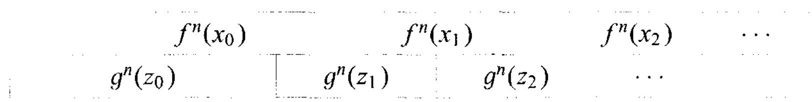
Next, let us demonstrate the following lemma:

Lemma 21. *Under the above-mentioned hypotheses, if* $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$ *and* x *is not ultimately periodic, then for any integer* n*, there exist letters* $\begin{bmatrix} u_n \\ v_n \end{bmatrix}, \begin{bmatrix} w_n \\ t_n \end{bmatrix}$ *in* B_q *and* a_n, b_n *in* A *such that* $u_n a_n, v_n a_n \in \mathrm{Fact}(\mathsf{x})$, $w_n b_n, t_n b_n \in \mathrm{Fact}(\mathsf{y})$ *and*

$$\varphi^n \left(\begin{bmatrix} u_n \\ v_n \end{bmatrix} a_n \right) \not\approx \psi^n \left(\begin{bmatrix} w_n \\ t_n \end{bmatrix} b_n \right).$$

Proof. Let $\mathsf{x} = x_0 x_1 \cdots x_p \cdots$. Let n be fixed. By Proposition 12, we can 'tile' x with $g^n(a)$, $a \in A$, that is to say

$$\exists \mathsf{z} = z_0 \cdots z_p \cdots, \ \exists r \text{ such that } \mathsf{x} = \sigma^r(g^n(\mathsf{z})) \text{ and } \mathrm{Fact}(\mathsf{z}) \subseteq \mathrm{Fact}(\mathsf{y}),$$



Now, we need two additional definitions.

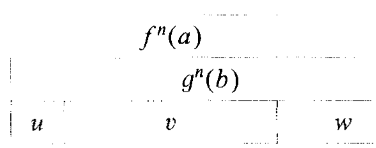
Definition 22. A *configuration of order* n is a 7-tuple $c = (a, b, u, \varepsilon, v, w, \eta)$ where $a, b \in A$, $u, v, w \in A^*$, $\varepsilon, \eta \in \{0, 1\}$ such that

$$f^n(a) = u^\varepsilon v w^\eta,$$

$$g^n(b) = u^{1-\varepsilon} v w^{1-\eta},$$

$$|v| > 0.$$

Example 23. Here $\varepsilon = \eta = 0$.



Definition 24. A configuration $c = (a, b, u, \varepsilon, v, w, \eta)$ has an *occurrence* at the point (i, j) if

$$|f^n(\mathsf{x}[0, i-1]) u^\varepsilon| = |\sigma^r g^n(\mathsf{x}[0, j-1]) u^{1-\varepsilon}|,$$

$$x_i = a,$$

$$z_j = b.$$

Let us go back to the proof of the proposition. There is a finite number of possible configurations (with n, f and g fixed), so there exists a configuration

$c = (\alpha, \beta, \delta, \varepsilon, \mu, v, \eta)$ which has two distinct occurrences (i, j) and (k, ℓ). So, we have

$$\delta^{1-\varepsilon} f^n(\mathsf{x}[i, +\infty[) = \delta^{\varepsilon} g^n(\mathsf{z}[j, +\infty[), \tag{6}$$

$$\delta^{1-\varepsilon} f^n(\mathsf{x}[k, +\infty[) = \delta^{\varepsilon} g^n(\mathsf{z}[\ell, +\infty[). \tag{7}$$

Furthermore, since x is not ultimately periodic, the sequences $x_i, x_{i+1}, \ldots$ and $x_k, x_{k+1}, \ldots$ cannot be equal, and it is the same for the sequences $z_j, z_{j+1}, \ldots$ and $z_\ell, z_{\ell+1}, \ldots$ Hence, there exist integers r and s such that

$$x_{i+r} \neq x_{k+r},$$

$$z_{j+s} \neq z_{\ell+s},$$

and which are minimal for these properties:

δ	$f^n(\mathsf{x}[i, i+r-1])$	$f^n(x_{i+r})$	$f^n(x_{i+r+1})$	$\cdots$
$g^n(\mathsf{z}[j, j+s-1])$		$g^n(z_{j+s})$	$g^n(z_{j+s+1})$	$\cdots$
$=$			$\neq$	
δ	$f^n(\mathsf{x}[k, k+r-1])$	$f^n(x_{k+r})$	$f^n(x_{k+r+1})$	$\cdots$
$g^n(\mathsf{z}[\ell, \ell+s-1])$		$g^n(z_{\ell+s})$	$g^n(z_{\ell+s+1})$	$\cdots$

Let $\pi = \delta^{1-\varepsilon} f^n(\mathsf{x}[i, i+r-1]) = \delta^{1-\varepsilon} f^n(\mathsf{x}[k, k+r-1])$ and $\rho = \delta^{\varepsilon} g^n(\mathsf{z}[j, j+s-1]) = \delta^{\varepsilon} g^n(\mathsf{z}[\ell, \ell+s-1])$.

Since f^n and g^n have bounded delay $2p-1$, there exist letters $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right], \left[\begin{smallmatrix} c' \\ d' \end{smallmatrix}\right] \in B_q$ and words u, v, w, u', v', w' such that

$$f^n(\mathsf{x}[i+r, i+r+q-1]) = ucv,$$

$$f^n(\mathsf{x}[k+r, k+r+q-1]) = udw,$$

and

$$g^n(\mathsf{x}[j+s, j+s+q-1]) = u'c'v',$$

$$g^n(\mathsf{x}[\ell+s, \ell+s+q-1]) = u'd'w',$$

π		u	c	v	$\cdots$
δ	$f^n(\mathsf{x}[i, i+r-1])$	$f^n(x_{i+r})$	$f^n(x_{i+r+1})$		$\cdots$
$g^n(\mathsf{z}[j, j+s-1])$		$g^n(z_{j+s})$	$g^n(z_{j+s+1})$		$\cdots$
ρ		u'	c'	v'	$\cdots$
$=$			$\neq$		
π		u	d	w	$\cdots$
δ	$f^n(\mathsf{x}[k, k+r-1])$	$f^n(x_{k+r})$	$f^n(x_{k+r+1})$		$\cdots$
$g^n(\mathsf{z}[\ell, \ell+s-1])$		$g^n(z_{\ell+s})$	$g^n(z_{\ell+s+1})$		$\cdots$
ρ		u'	d'	w'	$\cdots$

Equalities (6) and (7) give

$$\pi u c v f^n(x_{i+r+q}) \Leftrightarrow \rho u' c' v' g^n(z_{j+s+q}),$$

$$\pi u d w f^n(x_{k+r+q}) \Leftrightarrow \rho u' d' w' g^n(z_{\ell+s+q}).$$

Hence, we deduce that $\pi u = \rho u'$, $c = c', d = d'$ and

$$c v f^n(x_{i+r+q}) \Leftrightarrow c' v' g^n(z_{j+s+q}),$$

$$d w f^n(x_{k+r+q}) \Leftrightarrow d' w' g^n(z_{\ell+s+q}),$$

which can also be written

$$f^n(\mathsf{x}[i+r, i+r+q])/_q f^n(\mathsf{x}[k+r, k+r+q])$$
$$= g^n(\mathsf{z}[j+s, j+s+q])/_q g^n(\mathsf{z}[\ell+s, \ell+s+q]).$$

So, by Lemma 16

$$\varphi^n\left(\begin{bmatrix}\mathsf{x}[i+r, i+r+q-1]\\ \mathsf{x}[k+r, k+r+q-1]\end{bmatrix} x_{i+r+q}\right) \Leftrightarrow \psi^n\left(\begin{bmatrix}\mathsf{z}[j+s, j+s+q-1]\\ \mathsf{z}[\ell+s, \ell+s+q-1]\end{bmatrix} z_{j+s+q}\right).$$

Now, let

$$u_n = \mathsf{x}[i+r, i+r+q-1], \qquad v_n = \mathsf{x}[k+r, k+r+q-1],$$

$$w_n = \mathsf{z}[j+s, j+s+q-1], \qquad t_n = \mathsf{z}[\ell+s, \ell+s+q-1],$$

$$a_n = x_{i+r+q} \quad \text{and} \quad b_n = z_{j+s+q}.$$

We obtain the formula

$$\varphi^n\left(\begin{bmatrix}u_n\\ v_n\end{bmatrix} a_n\right) \Leftrightarrow \psi^n\left(\begin{bmatrix}w_n\\ t_n\end{bmatrix} b_n\right)$$

which completes the proof. □

Proof of Theorem 19 (Conclusion). Since there is a finite number of possibilities in the choice of the u_n, v_n, w_n, t_n, a_n and b_n in Lemma 21, there exist an infinite set of integers I and letters $\left[\begin{smallmatrix}u\\ v\end{smallmatrix}\right], \left[\begin{smallmatrix}w\\ t\end{smallmatrix}\right] \in B_q$, $a, b \in A$ such that

$$\forall n \in I, \quad \varphi^n\left(\begin{bmatrix}u\\ v\end{bmatrix} a\right) \Leftrightarrow \psi^n\left(\begin{bmatrix}w\\ t\end{bmatrix} b\right). \tag{8}$$

Let

$$F : \mathbb{N} \to B_q \times A,$$
$$n \mapsto \mathrm{Pref}_2\left(\varphi^n\left(\left[\begin{smallmatrix}u\\ v\end{smallmatrix}\right] a\right)\right),$$

$$G : \mathbb{N} \to B_q \times A,$$
$$n \mapsto \mathrm{Pref}_2\left(\psi^n\left(\begin{bmatrix} u \\ v \end{bmatrix} a\right)\right),$$

$$\theta : B_q \times A \to B_q \times A,$$
$$\begin{bmatrix} x \\ y \end{bmatrix} z \mapsto \mathrm{Pref}_2\left(\varphi^n\left(\begin{bmatrix} x \\ y \end{bmatrix} z\right)\right).$$

With these notations, we can write Eq. (8) as follows:

$$\forall n \in I, \quad F(n) = G(n). \tag{9}$$

Moreover, by definition of F and of θ, we have

$$\forall n \in \mathbb{N}, \quad F(n+1) = \theta(F(n)). \tag{10}$$

Since $B_q \times A$ is finite and I is infinite, there exist n_0 and p, with $p > 0$ such that $n_0 \in I$, $n_0 + p \in I$ and $F(n_0) = F(n_0 + p)$, which implies, by (10), that F is periodic of period p from n_0. For symmetric reasons, G too is periodic of period p from n_0 (notice that Eq. (9) implies $G(n_0) = F(n_0) = F(n_0 + p) = G(n_0 + p)$).

Since I is infinite, there exists i_0 such that the set $I_0 = \{n \in I \mid n - n_0 = i_0 \pmod p,\ n \geqslant n_0\}$ is infinite. Let $m_0 = \min(I_0)$ and let $J = \{n \in \mathbb{N} \mid m_0 + np \in I_0\}$. Since I_0 is infinite, the set J is infinite too. Let $\begin{bmatrix} \mu \\ \nu \end{bmatrix} \in B_q$ and $\alpha \in A$ be such that $F(m_0) = \begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha$. Then, for any integer n, one has $F(m_0 + np) = \begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha = G(m_0 + np)$ (by the periodicity of F and G). If, in addition, $n \in J$,

$$\varphi^{np}\left(\begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha\right) = \varphi^{np}\left(\mathrm{Pref}_2\left(\varphi^{m_0}\left(\begin{bmatrix} u \\ v \end{bmatrix} a\right)\right)\right) \leqslant \varphi^{np+m_0}\left(\begin{bmatrix} u \\ v \end{bmatrix} a\right).$$

Likewise,

$$\psi^{np}\left(\begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha\right) \leqslant \psi^{np+m_0}\left(\begin{bmatrix} w \\ t \end{bmatrix} b\right).$$

Therefore, Eq. (8) allows us to write

$$\varphi^{np}\left(\begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha\right) \not\asymp \psi^{np}\left(\begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha\right) \quad (n \in J).$$

Let φ' and ψ' be the endomorphisms on the alphabet $A \cup \{S\}$, where S is a new letter, defined as follows:

$$\varphi' : \begin{array}{l} S \mapsto Ss \qquad\qquad \left(\text{where } \varphi^p\left(\begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha\right) = \begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha s\right), \\ x \mapsto \varphi^p(x) = f^p(x) \quad (x \in A), \end{array}$$

and

$$\psi' : \begin{array}{l} S \mapsto Ss' \qquad\qquad \left(\text{where } \psi^p\left(\begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha\right) = \begin{bmatrix} \mu \\ \nu \end{bmatrix} \alpha s'\right) \\ x \mapsto \psi^p(x) = g^p(x) \quad (x \in A). \end{array}$$

Then for any $n \in J$,

$$\varphi'^n(S) \not\approx \psi'^n(S)$$

since φ' and ψ' are everywhere growing and J is infinite, we obtain that

$$\varphi'^\omega(S) = \psi'^\omega(S).$$

We choose p minimal, as in the definition of $\mathscr{C}$, because it does not change the infinite words $\varphi'^\omega(S)$ and $\psi'^\omega(S)$. Let us also notice that $\mu\alpha \in \mathrm{Fact}(\mathsf{y})$.

The existence of μ, v and α implies $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$. This follows from

$$\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\varphi'^\omega(S)) \cap A^* = \mathrm{Fact}(\psi'^\omega(S)) \cap A^* \subseteq \mathrm{Fact}(\mathsf{y}).$$

Let us prove the first inclusion $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\varphi'^\omega(S)) \cap A^*$. Let u be a factor of x. The morphism f is primitive; then there exists an integer n such that

$$u \in \mathrm{Fact}(\sigma(f^{np}(\alpha))) = \mathrm{Fact}(\sigma(\varphi^{np}(\alpha))) \subseteq \mathrm{Fact}(\varphi'^n(\alpha)) \cap A^*,$$

which indeed gives the required inclusion.

The second inclusion $\mathrm{Fact}(\psi'^\omega(S)) \cap A^* \subseteq \mathrm{Fact}(\mathsf{y})$ results from the fact that

$$\begin{aligned}\mathrm{Fact}(\psi'^\omega(S)) \cap A^* &= \mathrm{Fact}(\sigma^2(\psi^\omega(\mu\alpha))) \\ &\subseteq \mathrm{Fact}(\mathsf{y}) \quad (\text{because } \mu\alpha \in \mathrm{Fact}(\mathsf{y})). \quad \square\end{aligned}$$

5. Case of the binary alphabet

In this section, we prove that the subword inclusion problem is decidable for any pair of morphic words in a binary alphabet.

Throughout this section, the symbol A will denote the binary alphabet $\{a, b\}$. We recall that in this paper we consider only nonerasing morphisms.

First, we give a few preliminary results about the periodicity. The decidability of the periodicity has been solved in an effective way by Séébold.

Theorem 25 (Séébold [12]). *Let x be a morphic word generated by a (nonerasing) morphism f such that $\mathsf{x} = f^\omega(a)$. If x is periodic, then f has one of the five following forms*:

(i) $f(a) = a^p$, $p \geqslant 2$ *and* $f(b) \in A^*$ *and then* $\mathsf{x} = a^\omega$.
(ii) $f(a) = ab^p$, $f(b) = b^q$, $p, q \geqslant 1$ *and then* $\mathsf{x} = ab^\omega$.
(iii) $f(a) = (ab^p)^q a$, $f(b) = b$, $p, q \geqslant 1$ *and then* $\mathsf{x} = ((ab)^p)^\omega$.
(iv) $f(a) = v^p$, $f(b) = v^q$, $p, q \geqslant 1, |v| > 1$ *and then* $\mathsf{x} = v^\omega$.
(v) $f(a) = (ab)^p a$, $f(b) = (ba)^q b$, $p, q \geqslant 1$ *and then* $\mathsf{x} = (ab)^\omega$.

Remark 26. There is actually a sixth case if we assume that the morphism f may be erasing.

We are going to consider now the nonperiodic words. A finite word u is *primitive* if it is not a power of another word that is if

$$u = v^n, \quad n \geqslant 1 \Rightarrow n = 1.$$

Lemma 27. *Let* x *be a morphic word over the alphabet* A *generated by a morphism* f *such that* $\mathsf{x} = f^\omega(a)$ *and assume that* x *is not ultimately periodic. There exists a primitive word* u*, such that for any integer* n*, the word* u^n *is a factor of* x *if and only if* $u = b$ *and* f *has one of the two following forms*:

(i) $f(a) \in aA^*bA^* \setminus ab^*$ *and* $f(b) = b^p$ *for some* $p \geqslant 2$.

(ii) $f(a) \in aA^*b \setminus ab^*$ *and* $f(b) = b$.

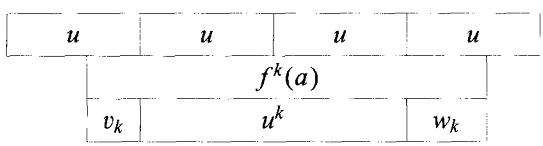
In these cases, for any integer n*, the word* ab^n *is also a factor of* x.

Proof. Let us assume that such a word u exists. In order that u exists, i.e. in order that x is strongly repetitive, the morphism f must not be primitive, i.e. $f(a) \in aA^*bA^*$ and $f(b) = b^p$ for some $p \geqslant 1$.

At first, let us show that necessarily $u = b$. To this end, let us assume that u is different from b. Let k_0 be an integer such that $|f^{k_0}(a)| > 3|u|$. For any integer $k \geqslant k_0$, there exists an integer n such that $|u^n| \geqslant 3\max\{f^k(a), f^k(b)\}$. Since u^n is a factor of x, there exist letters $x_1, \ldots, x_m$, $m \geqslant 3$ and integers r, s such that

$$u^n = \mathrm{Suff}_r(f^k(x_1))f^k(x_2)\ldots f^k(x_{m-1})\mathrm{Pref}_s(f^k(x_m)).$$

So there exist an integer $\ell_k \geqslant 1$, a suffix v_k of u and a prefix w_k of u such that $f^k(x_2) = v_k u^{\ell_k} w_k$. Since u contains the letter a, the letter x_2 cannot be equal to b:



So we have proved that, for any integer $k \geqslant k_0$, there exist an integer $\ell_k \geqslant 1$ and suffixes v_k of u and prefixes w_k of u such that $f^k(a) = v_k u^{\ell_k} w_k$.

Since u is primitive, it is easy to see that there exists a suffix v of u such that for any k, $v_k = v$. Therefore, $f^\omega(a) = vu^\omega$, so, since v is prefix of u, $\mathsf{x} = f^\omega(a)$ is periodic, which is in contradiction with the hypothesis. We have indeed proved that $u = b$.

It is clear that in the two cases of the lemma, the word ab^n is a factor of x for any integer n. Conversely, let us assume now that $f(a) \in aA^*a$ and $f(b) = b$. Let $n = |f(a)| + 1$. It is easy to see by induction that for any integer k, the word b^n is not a factor of $f^k(a)$. Indeed, in order that b^n is a factor of $f^{k+1}(a)$, it must be a factor of $f^k(a)$.

The two cases mentioned above are indeed the only possible ones. $\square$

Most of the following notations and definitions are from [4]. An endomorphism f on A^* is said to be *marked* if $\mathrm{Pref}_1(f(a)) \neq \mathrm{Pref}_1(f(b))$. It is *well marked* if

$\mathrm{Pref}_1(f(a)) = a$ and $\mathrm{Pref}_1(f(b)) = b$. We define the following function $\mathrm{cyc}_1 : A^* \to A^*$ by

$$\mathrm{cyc}_1(1) = 1,$$

$$\mathrm{cyc}_1(xu) = ux \quad \text{if } x \in A \text{ and } u \in A^*,$$

which can also be written $\mathrm{cyc}_1(u) = \sigma(u)\mathrm{Pref}_1(u)$. Let $\mathrm{cyc}_n = (\mathrm{cyc})^n$ for $n \geqslant 1$. Therefore, for any word u over A, we have

$$\mathrm{cyc}_n(u) = (\mathrm{Pref}_r(u))^{-1} \cdot u \cdot \mathrm{Pref}_r(u),$$

where r is the remainder of the Euclidean division of n by $|u|$.

It is known that $f(ab) = f(ba)$ if and only if there exist a word u and integers p and q such that $f(a) = u^p$ and $f(b) = u^q$. In particular, x is therefore periodic. Let us assume, now, that $f(ab) \neq f(ba)$. Let π_f the longest common prefix of $f(ab)$ and $f(ba)$. Then the function $f' : A^* \to A^*$ defined by $f' = \mathrm{cyc}_{|\pi_f|} \circ f$ is a marked morphism. It is then obvious that f'^2 is a well-marked morphism. This morphism will be denoted by $M(f)$.

Lemma 28. *Let f be an endomorphism on A^* which generates an infinite word $\mathsf{x} = f^\omega(a)$. Then, x is not recurrent if and only if there exist positive integers n and m such that*

$$f(a) = ab^n, \qquad f(b) = b^m.$$

In this case, $\mathsf{x} = ab^\omega$.

Proof. If f is primitive, then x has bounded gaps and then it is recurrent. If $f(a) \in aA^*aA^*$, then x is recurrent too. If we are not in one of these above-mentioned two cases, then f has the form

$$f(a) = ab^n, \qquad f(b) = b^m.$$

In this case $\mathsf{x} = ab^\omega$ which is clearly not recurrent. □

Lemma 29. *Let x be a nonultimately periodic morphic word generated by a morphism f such that $\mathsf{x} = f^\omega(a)$. Let $\mathsf{x}' = (M(f))^\omega(a)$. Then the equality $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{x}')$ is satisfied.*

Proof. Since x is not ultimately periodic, x is recurrent (cf. Lemma 28). If x is not ultimately periodic, then $f(ab) \neq f(ba)$ and we define π_f and $M(f)$ as above. Let $f' = M(f)$.

On the one hand, $f(\mathsf{x}') = \pi_f \cdot f'(\mathsf{x}') = \pi_f \mathsf{x}'$. So $\mathrm{Fact}(f(\mathsf{x}')) \subseteq \mathrm{Fact}(f'(\mathsf{x}')) = \mathrm{Fact}(\mathsf{x}')$, and for any n, $\mathrm{Fact}(f^n(\mathsf{x}')) \subseteq \mathrm{Fact}(\mathsf{x}')$. Therefore, we have $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(f^\omega(a)) \subseteq \mathrm{Fact}(\mathsf{x}')$. On the other, $f'(\mathsf{x}) = \mathrm{cyc}_{|\pi_f|} \circ f(\mathsf{x}) = \pi_f^{-1}\mathsf{x}$. Since x is recurrent, this implies $\mathrm{Fact}(f'(\mathsf{x})) = \mathrm{Fact}(f(\mathsf{x})) = \mathrm{Fact}(\mathsf{x})$. In the same way as above, we can deduce that $\mathrm{Fact}(\mathsf{x}') = \mathrm{Fact}(f'^{\omega}(a)) \subseteq \mathrm{Fact}(\mathsf{x})$. We indeed obtain the stated equality. □

Lemma 30. *Let* x *and* y *be two morphic words, respectively, generated by the well-marked morphisms* f *and* g *on* A^*. *If* x *is not ultimately periodic and if* $\mathrm{Fact}(\mathsf{x}) = \mathrm{Fact}(\mathsf{y})$, *then for any integer* n,

$$f^n(a) \not\approx g^n(a), \qquad f^n(b) \not\approx g^n(b).$$

In addition, for all integers p, n *and* m, *there exist words* $u, u', v, v' \in A^*$, $|u| = |u'| = n$, $|v| = |v'| = m$ *such that* $au, bu' \in \mathrm{Fact}(\mathsf{x})$, $av, bv' \in \mathrm{Fact}(\mathsf{y})$ *and*

$$f^p(au) \not\approx g^p(av), \qquad f^p(bu') \not\approx g^p(bv').$$

Proof. We only have to notice that if f is well marked, for any integer q of the form $4r - 3$, then

$$\forall n, \forall \begin{bmatrix} au \\ bv \end{bmatrix} \in B_q, \qquad \varphi^n \begin{bmatrix} au \\ bv \end{bmatrix} = f^n(au),$$

where $\varphi = T_q(f)$, and so is for g. It remains to use Lemma 21 noticing that the fact that f and g are prefixes makes the hypothesis that f and g are everywhere growing useless. □

Now, we can state the theorem.

Theorem 31. *Let* x *and* y *be two morphic words over the alphabet* $A = \{a, b\}$. *Then the inclusion* $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$ *is decidable.*

Proof. First, let us give the algorithm of decidability which we shall use.

Algorithm 3. At first, we test whether x is ultimately periodic by means of Theorem 25.

(i) If x is ultimately periodic, we test whether y is ultimately periodic.
   (a) If y is not ultimately periodic, then we have two cases if neither b nor a are period of x, then in view of Lemma 27 we cannot have $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$. If b or a is period of x, then in view of Theorem 25, four cases are possible: $\mathsf{x} = b^{\omega}$, $\mathsf{x} = ab^{\omega}$, $\mathsf{x} = a^{\omega}$ or $\mathsf{x} = ba^{\omega}$ and we can test by means of Lemma 27 whether all the factors of x are factors of y.
   (b) Otherwise, if y is ultimately periodic: we can compute the periods and the preperiods of x and y by means of Theorem 25 and therefore we can easily decide whether $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$.

(ii) If neither x nor y is ultimately periodic, we describe below how we can decide the inclusion $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$.

Decidability of case (ii): We consider here two morphic words x and y nonultimately periodic. Let f and g be two morphisms generating, respectively, x and y. According to Lemma 29, we can assume that f and g are well marked without changing the sets $\mathrm{Fact}(\mathsf{x})$ and $\mathrm{Fact}(\mathsf{y})$. In order to simplify let us assume that $\mathsf{x} = f^{\omega}(a)$. We have three cases.

$g(a) \neq a$: Then $g^{\omega}(a)$ is an infinite word. Let us assume that $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$. By Lemma 30, for any integer n, we have $f^n(a) \not\Leftrightarrow g^n(a)$, which implies $f^{\omega}(a) = g^{\omega}(a)$.

Conversely, if $\mathsf{x} = f^{\omega}(a) = g^{\omega}(a)$, one has $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$. We have then to test $f^{\omega}(a) = g^{\omega}(a)$ in order to decide the inclusion. $g(a) = a$ and $f(b) \neq b$: Then $\mathsf{y} = g^{\omega}(b)$. Let us assume that $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$. Therefore, by the same reasoning as above $f^{\omega}(b) = g^{\omega}(b)$. Conversely, if $f^{\omega}(b) = g^{\omega}(b)$ then, since $\mathsf{y} = g^{\omega}(b)$ is not ultimately periodic $f^{\omega}(b) \neq b^{\omega}$, hence $f(b) \in bA^*aA^*$, which implies that f is primitive. We can then deduce the inclusion $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$. We then have to test $f^{\omega}(b) = g^{\omega}(b)$.

$g(a) = a$ and $f(b) = b$. Then $\mathsf{y} = g^{\omega}(b)$.

Let us assume that $\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\mathsf{y})$. Since x is not ultimately periodic, there exist a positive integer n and a word u such that

$$f(a) = a^n bu, \qquad f(b) = b.$$

There exists too a nonnull word v such that

$$g(a) = a, \qquad g(b) = bv.$$

Then, for any integer p, the word $a^n b$ is a prefix of $f^p(a)$. But, according to Lemma 30, for all integer p, there exists a word w_p of length n such that $f^p(a) \not\Leftrightarrow g^p(aw_p)$. Since g is well marked and $g(a) = a$, we must have $w_p = a^{n-1}b$. So, we have

$$\forall p, \quad f^p(a) \not\Leftrightarrow g^p(a^n b),$$

which implies

$$f^{\omega}(a) = g^{\omega}(a^n b)$$

and also

$$f^{\omega}(f(a)) = f^{\omega}(a^n bu) = g^{\omega}(a^n b) = a^n g^{\omega}(b). \tag{11}$$

Let u' be the word such that $f(a^n b) = a^n bu'$. The word u' is not the null word. Now, if we define the two morphisms φ and ψ on $A \cup \{S\}$, where S is a new letter, by

$$\varphi : \begin{array}{l} S \mapsto Su' \\ a \mapsto f(a) \\ b \mapsto f(b), \end{array}$$

and

$$\begin{aligned} &S \mapsto Sv \\ \Psi :\ &a \mapsto g(a) \\ &b \mapsto g(b). \end{aligned}$$

Then for any integer p, we have

$$\varphi^p(S) = S \cdot (a^n b)^{-1} f^p(a^n b),$$
$$\psi^p(S) = S \cdot (b)^{-1} g^p(b).$$

Then, Eq. (11) implies

$$\varphi^\omega(S) = \psi^\omega(S).$$

Conversely, we can always define φ and ψ as above. Let us assume that $\varphi^\omega(S) = \psi^\omega(S)$. Since, by hypothesis, x is not ultimately periodic, it is recurrent (cf. Lemma 28). It is not difficult to see that we indeed have the inclusion

$$\mathrm{Fact}(\mathsf{x}) \subseteq \mathrm{Fact}(\varphi^\omega(S)) \cap A^* = \mathrm{Fact}(\psi^\omega(S)) \cap A^* \subseteq \mathrm{Fact}(\mathsf{y}). \quad \square$$

Acknowledgements

I would like to thank J. Berstel who proposed this problem to me and for fruitful discussions, J. Cassaigne for suggesting me a generalization of a previous version of Proposition 12 and P. Séébold. I would also like to thank the referee for his useful remarks.

References

[1] J. Berstel and P. Séébold, A remark on morphic Sturmian words, RAIRO Inform. Theoret. Appl. 28 (1994) 225–263.

[2] A. Cobham, Uniform tag sequences, Math. Systems Theory 6 (1972) 164–192.

[3] K. Čulik II and I. Fris, The decidability of the equivalence problem for D0L-systems, Inform. Control 35 (1977) 20–39.

[4] K. Čulik II and T. Harju, The ω-sequence equivalence problem for D0L systems is decidable, J. Assoc. Comput. Mach. 31 (1984) 282–298.

[5] A. Ehrenfeucht, J. Karhumäki and G. Rozenberg, On binary equality sets and a solution to the test set conjecture in the binary case, J. Algebra. 85 (1983) 76–85.

[6] A. Ehrenfeucht and G. Rozenberg, Repetition of subwords in D0L languages, Inform. Control 59 (1983) 13–35.

[7] T. Harju and M. Linna, On the periodicity of morphisms on free monoids, RAIRO Inform. Theoret. Appl. 20 (1986) 47–54.

[8] B. Mossé, Puissance de mots et reconnaissabilité des points fixes d'une substitution, Theoret. Comput. Sci. 99 (1992) 327–334.

[9] J. Pansiot, Decidability of periodicity for infinite words, RAIRO Inform. Theoret. Appl. 20 (1986) 43–46.
[10] M. Queffelec, Substitution Dynamical Systems-Spectral Analysis, Lecture Notes in Math., Vol. 1294 (Springer, Berlin, 1987).
[11] A. Salomaa, Jewels of Formal Language Theory (Computer Science Press, Rockville, MD, 1981).
[12] P. Séébold, An effective solution to the D0L periodicity problem in the binary case, EATCS Bull. 36 (1988) 137–151.